Deep Multitask Learning for Ultrasound Beamforming

Elay Dahan

Deep Multitask Learning for Ultrasound Beamforming

Research Thesis

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering

Elay Dahan

Submitted to the Senate
of the Technion — Israel Institute of Technology
Shvat 5785 Haifa February 2025

This research was carried out under the supervision of Prof. Israel Cohen in the Faculty of Electrical and Computer Engineering.

Some results in this thesis have been published in journals during the course of the author's research period.

The author of this thesis states that the research, including the collection, processing, and presentation of data, addressing and comparing to previous research, etc., was done entirely in an honest way, as expected from scientific research that is conducted according to the ethical standards of the academic world. Also, reporting the research and its results in this thesis was done in an honest and complete manner, according to the same standards.

Acknowledgements

I would like to extend my thanks to my research supervisor, Prof. Israel Cohen, without whose support and contribution this research would not have been possible. Throughout this project, he guided me on how to conduct research and equipped me with valuable skills that helped me overcome difficulties and gaps in knowledge in certain areas. His mentorship was instrumental in my development during this research endeavor. Additionally, I would like to thank my colleagues for their consultations and contribution to this work.

I would also like to thank my parents Haim and Sigal and my partner Noa. Without their support throughout this journey this research would not be possible.

The generous financial help of the Technion is gratefully acknowledged.

List of Publications

1. Elay Dahan and Israel Cohen, "Deep-learning-based multitask ultrasound beamforming," *Information*, Special Issue on Deep Learning for Image, Video and Signal Processing, Vol. 14, No. 10, Article 582, pp. 1-15, October 2023.

2. Elay Dahan and Israel Cohen, "Lightweight Multitask Deep Learning for Ultrasound Image Formation," submitted.

Contents

List of Figures

Abstract

In this thesis, we investigate the challenges of beamformer design for multitask ultrasound beamforming. Beamforming is a critical technique for enhancing signal quality from specific directions while suppressing interference from others. It is widely utilized in medical imaging, especially ultrasound, significantly impacting image resolution and clarity.

This thesis addresses two primary problems. The first problem focuses on developing a weight normalization scheme for multitask learning in ultrasound beamforming. The objective is optimizing beamforming for high frame rate requirements and image denoising. By fitting weight normalization parameters for sub-sampling tasks and optimizing them for speckle reduction, our fully convolutional neural network outperforms traditional single-angle delay-and-sum methods on pixel-level metrics, improving speckle noise reduction, subsampling, and single-angle reconstruction. The second problem explores a lightweight, multitask-based deep learning approach for ultrasound image formation. This approach aims to enhance image resolution and contrast, which are typically lower in ultrasound than other imaging modalities. We introduce a novel method that learns a feature transformation from the beamforming domain to speckle reduction parameter-efficiently, ensuring low computational costs. Qualitative evaluations demonstrate that this method achieves performance comparable to dedicated network training while significantly outperforming single-angle delay-and-sum techniques in speckle noise reduction.

First, we present the weight normalization scheme implemented in a convolutional neural network for multitask ultrasound beamforming. This method adapts to dynamic imaging environments by optimizing weight normalization parameters for various tasks, ensuring high frame rates and effective speckle noise reduction. Extensive simulations and experiments validate the superior performance of this approach compared to conventional methods. We began by benchmarking our model for ultrasound image formation using the deep learning ultrasound image formation challenge. Our results were compared to those of other challenge participants, and our approach achieved the best results across all global and local image quality metrics. Subsequently, we tested our multitask approach on two different tasks: speckle reduction and reconstruction from sub-sampled signals. Once again, our model outperformed the conventional DAS with post-processing in every global and local image quality metric, demonstrating the

effectiveness of our method for multitask learning.

Next, we introduce an efficient algorithm for lightweight multitask-based deep learning in ultrasound image formation. This algorithm optimizes the feature transformation process, maintaining high performance with minimal computational overhead. Our design achieves higher resolution and contrast in ultrasound images, addressing the typical limitations of lower-frequency imaging. Finally, we provided a theoretical analysis of the parameter overhead required by this lightweight approach, demonstrating that the overhead for transforming knowledge for additional tasks is less than 1% of the model parameter count. We conclude that the proposed approach is suitable for practical applications in ultrasound image formation, as it provides scalability, efficiency, and high image quality.

Abbreviations

AdaIN	:	Adaptive Instance Normalization
B-Mode	:	Brightness Mode
CNN	:	Convolutional neural network
CNR	:	Contrast-to-noise ratio
CT	:	Computed Tomography
CUBDL	:	Challenge on Ultrasound Beamforming with Deep Learning
DAS	:	Delay and sum
DICE	:	Sørensen–Dice coefficient
DNN	:	Deep Neural Network
F-DMAS	:	Filter Delay-Multiply and Sum beamforming algorithm
FWHM	:	Full Width at Half Maximum
gCNR	:	Generalized contrast-to-noise ratio
GELU	:	Gaussian Error Linear Unit
IQ	:	In-phase, Quadrature
LSTM	:	Long Short-Term Memory
MRI	:	Magnetic Resonance Imaging
MS-SSIM	:	Multi-Scale Structural Similarity
MVDR	:	Minimum variance distortionless response
PSNR	:	Peak Signal-to-Noise Ratio
ReLU	:	Rectified Linear Unit
RF	:	Radio Frequency
ROI	:	Region of Interest
SGD	:	Stochastic Gradient Descent
SNR	:	Signal-to-Noise Ratio
SSIM	:	Structural Similarity Index Measure

Notations

$b_{i,k}$	:	Bias parameter for layer i and task k in the modified convolution weights
C	:	Number of channels
E	:	Number of transmit events
f_i	:	Histogram of pixel values of the i'th region of Interest
$f(X, k; \theta)$	:	Neural network function with input X, task index k, and model parameters θ
I	:	Final image after envelope detection and log compression
$\hat{I}$	:	Estimated multi-angle reconstruction
$\mathrm{IQ}_{\mathrm{tof}} \in \mathbb{C}^{E \times C \times N_x \times N_y}$	:	IQ tensor after time-of-flight correction
K_1	:	Dimension of the kernel (width)
K_2	:	Dimension of the kernel (height)
L	:	Loss Function for neural network training
N_t	:	Number of samples in each received signal
N_x	:	Spatial dimension after upsampling process (width)
N_y	:	Spatial dimension after upsampling process (height)
$\mathbf{R_x}$	:	Received signal covariance matrix in the MVDR beamformer optimization problem
$s_{i,k}$	:	Scale parameter for layer i and task k in the modified convolution weights
T	:	Transformation matrix for linear transformation
T_{ij}	:	Transformation matrix for layer i and task j
$\mathbf{w}$	:	Apodization weights in the MVDR beamformer optimization problem
W	:	Apodization tensor used in model-based beamforming algorithms
W	:	RF data Window length
W_i	:	Learned filter weights for the i'th convolutional layer
W_{ij}	:	Transformed matrix at layer i for task j
X	:	Received radio frequency (RF) signal
x_n	:	Envelope of the beamformed reconstruction
Y	:	Aggregated signal after beamforming

y_n	:	Ground-truth envelope
α	:	Parameter controlling the intensity of the task-specific effect in the modified convolution weights
β_1, β_2	:	AdamW optimizer hyper parameters
ℓ_1	:	Mean Absolute Error
ℓ_2	:	Mean Squared Error
μ_i	:	Mean pixel value of the i'th region of Interest
ρ	:	Normalized cross-correlation
σ_i	:	Standard deviation of the pixel values of the i'th region of Interest
$\tau_{c,e,r}$	:	Time delay for each pixel in each transmit event, calculated based on the geometry of the transducer
$\mathbb{C}^{E \times C \times N_t}$	:	Complex tensor representation of IQ signals

Chapter 1

Introduction

1.1 Background and Motivation

Ultrasound stands out as one of the most widely used medical imaging techniques. Unlike magnetic resonance imaging (MRI) or computed tomography (CT), ultrasound devices are relatively compact and cost-effective. These various modalities enable medical professionals to explore the internal structures of a patient's body, offering supplementary insights into the patient's condition for physicians.

In certain situations, like emergency rooms, precise and quick diagnosis is essential. Because ultrasound imaging is the fastest, most compact, and portable among other imaging methods, it's an ideal choice for such scenarios. Moreover, ultrasound waves are non-ionizing, unlike CT and X-ray scans, making it a safe medical imaging tool. Compared to other imaging techniques like MRI and CT scans, ultrasound's primary drawback is its lower image quality. This is because the human body's inherent non-uniformity and the diverse composition of its tissues, each with unique physical properties, result in higher noise levels in the ultrasound waves captured during imaging. Additionally, the relatively lower frequency, and consequently higher wavelength of ultrasonic waves, in comparison to other imaging modalities, diminishes spatial resolution. Unclear images stemming from noise and decreased resolution can complicate the diagnostic process for medical professionals, potentially resulting in errors or inaccurate diagnoses. Therefore, producing high-quality ultrasound images is imperative for quick and precise diagnosis.

In modern practice, employing image denoising and enhancement algorithms is commonplace after generating the ultrasound image. Speckle noise reduction [MF09] is one such example. In the context of ultrasound imaging, speckle noise has been studied extensively. There are different approaches for speckle noise reduction in ultrasound, for example, non-local filtering methods [CHKB09; SG17] and deep learning methods [LZ20; DSOL18; KNZL22]. Image denoising post-processing step reduces the device's frame rate, which is not optimal.

Commercial ultrasound devices incorporate hundreds of post-processing algorithms

and image optimizations. These various algorithms and configurations significantly increase the computational complexity of the reconstruction pipeline. Additionally, each model or configuration requires storing a set of relevant parameters, which becomes highly inefficient in practical cases where hundreds of such post-processing functions exist. This research focuses on proposing a novel multitask learning framework for ultrasound beamforming. Ideally, a single unified model can generate all target images, improving computational efficiency by eliminating sequential processing and reducing storage costs by consolidating all algorithms into a single model.

In addition to image denoising and enhancement, the medical imaging community is increasingly adopting automatic image analysis algorithms such as classification and segmentation. Similar to other computer vision challenges, deep learning methods offer state-of-the-art results for medical imaging tasks. However, neural networks are vulnerable to out-of-distribution data samples. Therefore, noisy images can result in incorrect predictions by the neural network. To maintain consistent and dependable performance of deep neural networks (DNNs) for image analysis tasks, it is crucial to ensure a certain level of quality in brightness mode (B-Mode) ultrasound images.

Ultrasound imaging operates by utilizing ultrasonic waves. These waves are emitted by the transducer and upon receiving reflected echo signals, they are used to create the resulting image. The transducer records the reflected wave, encoding each signal into a pixel value. The grayscale value of each pixel is determined by the properties of the reflected signal. A lower received signal power, in comparison to the transmitted energy, indicates high ultrasonic wave absorption and is encoded to a lower grayscale value. Conversely, higher received signal power indicates low ultrasonic wave absorption and is encoded to higher pixel values.

The ultrasound transducer consists of N transmitters, with each transmit event involving a subset of transducer elements chosen to emit ultrasonic waves, which are then received by the receiving elements. The imaging scheme determines the different sets of receiving/transmitting elements. For instance, in focused transmission, each transmit event captures a depth-wise line within the target tissue, with each transmit element focused on the target line by applying an appropriate transmit time delay.

In focused transmission, also referred to as line scanning, the entire image reconstruction process is time-consuming as each line is acquired separately. However, when all transducer elements are utilized to transmit a plane-wave ultrasonic wave, the entire region is captured with a single transmission. This generated plane ultrasonic wave is transmitted at various angles within each transmission event. By applying specific time delays to each transmission element, the combination of these time-delayed excitation signals forms an angled plane wave. When employing a plane wave, the resulting echoes represent multiple lines at a single transmission event.

As a result, when considering the same penetration depth of ultrasonic waves, employing unfocused transmit leads to a higher frame rate compared to focused transmit. While plane-wave transmission offers faster and more suitable options for real-time

imaging, it is accompanied by reduced resolution and contrast compared to focused transmission. Therefore, the image reconstruction algorithm plays a critical role in ensuring optimal overall performance.

The formation of an ultrasound image entails the following steps:

1. Receiving the echo of the generated ultrasonic wave;
2. Applying time-of-flight correction to the received signal;
3. Beamforming the array of time-aligned signals;
4. Applying log compression;
5. Performing image post-processing.

In commercial ultrasound devices, a low-complexity delay and sum (DAS) beamforming algorithm is typically chosen to maintain a high frame rate. This approach involves using predetermined static delays for time-of-flight correction on received signals, followed by summation of channel data to generate a beamformed signal. However, the low computational complexity of DAS can compromise the overall width of the beamformed signal's main lobe and side lobes. More advanced adaptive algorithms, such as the minimum variance distortionless response (MVDR), offer improved performance. With adaptive beamforming, the summation weights are not constant but are calculated from the data, resulting in better results. Despite the superior performance of adaptive beamformers like MVDR compared to DAS, they face significant computational complexity challenges that make them unsuitable for real-time applications.

Deep learning has demonstrated remarkable achievements across diverse tasks, including image processing, speech recognition, and more. Particularly in medical imaging, deep learning has emerged as the leading approach, exhibiting state-of-the-art performance in tasks such as image classification and segmentation. For instance, Chen *et. al.*[XCZ+20] showcased significant success in cerebrovascular segmentation from time-of-flight MRI data. Their approach, set within the framework of semi-supervised learning, incorporated two identical neural networks—one trained on labeled data and the other on unlabeled data—sharing weights. They utilized cross-entropy loss for labeled data and introduced a consistency loss term for unlabeled data, ensuring consistency between the input data and a perturbation of the sample, thereby ensuring the same segmentation map for a given sample and its perturbation. Their model achieved state-of-the-art results in terms of the DICE score [CRG+20].

Deep learning strategies have been applied to improve the performance of model-based and data-adaptive approaches like DAS and MVDR in terms of computational performance and image quality. Incorporating a data-driven approach like deep learning can result in a reduction in computational performance. For example, estimation of the MVDR output image with a neural network can reduce the results. In [LCdB+20], the authors have shown that they were able to generate images on par in terms of perceptual

quality to MVDR while maintaining a computational complexity of $O(n^2)$ compared to $O(n^3)$. Additionally, with deep learning, one can combine multiple sequential steps of the image formation pipeline, like beamforming and denoising into a single faster neural network.

Goudarzi *et. al.* [GAR20] proposed a MobileNetV2 [SHZ+18], neural network to estimate the reconstruction of a multi-angle DAS beamformer from a single-angle acquisition. the network input is a $2 \times C \times W$ tensor, where C is the number of receiving channels and W is the spatial window size set to 32. The network output is a two-element vector representing the IQ elements of the estimated multi-angle DAS. With a parameter count of 2.226 million, MobileNetV2 is considered a relatively lightweight neural network, enabling faster computation and inference times. However, since the reconstruction is performed pixel-by-pixel, the performance does not meet the requirement for real-time applications.

Rothlübbers *et. al.* [RSE+20] adopted a distinct methodology wherein the direct estimation of multi-angle in-phase and quadrature (IQ) components was replaced. Instead, the output of the DNN was employed as the beamforming weights. The resultant weights were subsequently multiplied with the input from a single angle to form the multi-angle estimate. The training data are 107 samples of privately acquired raw ultrasound RF data and publicly available data [LRC+16]. The network is then trained with a linear combination of mean squared error per pixel loss and multi-scale structural similarity (MS-SSIM) [WSB03].

Following the beamforming and log compression of the received echo signal, the subsequent step in the ultrasound image pipeline is the post-processing step. The post-processing steps are usually applied to improve the contrast and reduce the noise of the beamformed signal. Noise reduction is particularly crucial in situations where the notified area is more expansive, as observed in the case of plane wave ultrasound transmission. This is due to the tendency of plane wave ultrasound to exhibit higher noise levels and lower spatial resolution compared to focused transmit ultrasound. In medical imaging, post-processing operations on images, such as noise reduction, automatic segmentation, and classification, hold significant value in automating the diagnostic process or enhancing image quality. Denoised images offer a higher level of clarity, thereby aiding the medical team in the diagnostic process. Denoised images improve accuracy and efficiency in medical diagnoses by providing a more distinct visualization of anatomical structures and pathological features.

A common approach nowadays is to apply a task-specific algorithm after the image has been formed. Applying additional subsequent algorithms after the image formation decreases the framerate, which is not optimal. Another issue with that approach is that for every new task, a new separate neural network has to be trained or explicitly designed for the required task. Integrating a beamforming algorithm that can reconstruct the post-processed beamformed image directly, without incorporating external algorithms or methods in addition to the beamforming process, holds substantial

significance. A single beamformer that also outputs a post-processed image offers notable benefits in terms of improved performance and enhanced stability in end-to-end performance.

Bhatt *et. al.* [BNKL20] proposed a UNet-based architecture [RFB15] to predict segmentation and image formation reconstruction. The proposed architecture is based on one encoder and two decoders. Each decoder outputs a task-specific output. The first one outputs an ultrasound image reconstruction, and the second one outputs a segmentation map. One significant advantage of that approach is that the model outputs both a segmentation map and an ultrasound image simultaneously. Also, using one single encoder allows the model to learn features relevant to both tasks and then decode the global features of each task by a separate encoder. A disadvantage of this approach is that the computational resources required for running this model grow proportionally to the number of desired tasks since each requires its encoder. Furthermore, a new encoder must be trained from scratch for each future task.

Khan *et. al.* [KHY21b] proposed a different approach; they trained a U-Net variation. To control the task-specific output, they added adaptive instance normalization layers [HB17] (AdaIN) at the bottleneck block of the U-Net architecture. In parallel to the primary U-Net beamformer, they also trained a small, fully connected neural network that maps a style code to the AdaIN parameters-normalization mean and variance. After which, a normalization with task-specific mean and variance is applied to the output of the bottleneck block. The advantages of the approach proposed by Khan *et. al.* [KHY21b] are:

1. Scalability: given enough task-specific data, one has to train only a small portion of their complete architecture—the fully connected AdaIN layer parameters;

2. Performance: during inference, the AdaIN parameters can be pre-computed, and hence only a single forward pass of the U-Net network is required to generate task-specific output.

With the approach proposed by Khan *et. al.* [KHY21b], there is an evident improvement in both scalability and performance. However, it is essential to note that, for each task, only the representation of the bottleneck layer is modified. As a result, the task-specific output is solely controlled by employing different normalization techniques on the output of the bottleneck layer.

1.2 Main Contributions

This thesis concentrates on addressing the gaps in knowledge discussed earlier. Here are our main contributions:

- We propose a novel architecture for ultrasound beamforming that enhances efficiency compared to existing proposals in the field. This architecture processes

image patches in a single forward pass, offering a streamlined approach. Our proposed method has demonstrated superior results in the ultrasound image formation challenge, excelling in both perceptual quality metrics and computational efficiency.

- We introduce a novel weight normalization layer tailored for convolutional neural networks, which we apply to multi-task learning within the realm of ultrasound beamforming. Our approach demonstrates the capability to transfer trained knowledge with minimal performance degradation. We tested our multitask learning approach on speckle noise reduction and reconstruction from sub sampled signal. we have shown that our method yielded results that are superior to those of classical reconstruction algorithms.

- Additionally, we introduce a robust linear feature transformation method derived from multitask learning, which demonstrates minimal or negligible performance degradation while achieving enhanced parameter efficiency and improved image quality. We tested this approach on speckle reduction task and compared it to the weight normalization layer, and neural network that has been trained specifically for speckle reduction. we have shown that this approach yielded results that are on par with the specifically trained neural network and outperform the weight normalization scheme.

1.3 Research Overview

This research focuses on developing switchable multitask ultrasound beamformers. We propose a novel deep neural network architecture tailored for ultrasound beamforming, integrating custom layers for multitasking. Initially, we introduce a fully convolutional neural network approach, enabling rapid reconstruction by generating patches of pixels in each forward pass. Our subsequent exploration delves into multitask learning using convolutional layers, where our proposed weight normalization layer demonstrates promising results in both sub-sample reconstruction and speckle reduction. Evaluating our method against the CUBDL challenge dataset, we compare favorably with the challenge winners, showcasing our lightweight approach and superior performance across most metrics.

In the second phase of our research, we introduce an even lighter multitask learning approach, particularly suited for ultrasound beamforming, tested on CUBDL. Here, we propose a lightweight linear feature transformation module operating on each convolutional layer. Our results indicate that this method efficiently reduces parameter overhead per task while maintaining performance comparable to task-specific networks. This underscores the effectiveness of multitask learning in transferring knowledge from the base task of ultrasound beamforming to the target task of speckle reduction, achieving this with minimal parameter usage and minor performance tradeoffs.

Throughout this research, we prioritize reconstruction time, crucial for meeting the real-world demand for high-framerate ultrasound image formation pipelines. To ensure compatibility with this requirement, we design our neural network to be lightweight, facilitating swift processing and enabling high frame rates. Furthermore, our multi-task approach is crafted with efficiency in mind, ensuring that storing and loading parameters on ultrasound scanners remains viable for commercial devices. Additionally, we validate our model on both simulated ultrasound data and real in-vivo data, demonstrating its effectiveness in real-life scenarios.

1.4 Organization

This thesis is organized as follows. Chapter 2 presents the problem of ultrasound image formation and specifically ultrasound beamforming. In this chapter, we introduce the notations used throughout this thesis. In Chapter 3 A novel method for deep multitask learning applied to ultrasound beamforming. We compare our method to participants in the CUBDL challenge. In Chapter 4, we present a more efficient method for multitask learning. we show that the proposed method shows results on par with a network trained specifically for speckle reduction tasks thus proving the effectiveness of our method. In Chapter 5 we conclude the thesis, summarize the main contributions, and discuss future research directions.

Chapter 2

Preliminaries

This chapter provides the scientific background to the problems tackled in this thesis and an overview of the mathematical models used.

2.1 Problem Formulation

In plane-wave imaging, each transducer element records the received echo signal. The resulting echo signal from the plane wave is represented as a tensor $X \in \mathbf{R}^{C \times E \times N_t}$, where C is the number of receiving channels, E is the number of transmit events, and N_t is the number of time samples recorded. Subsequently, time-of-flight correction is implemented on the received signal to ensure precise alignment in terms of timing. The time delays for the correction are determined based on the transducer geometry relative to each pixel.

The subsequent step in image formation involves beamforming the time-aligned signals to produce the final image $Y \in \mathbf{R}^{N_X \times N_Y}$. Beamforming is a signal processing technique used for sensor arrays to consolidate multiple sources into a unified signal. Since ultrasound transducers comprise multiple sensors, after emitting an ultrasonic wave from the transmit elements, the resulting echo is received by a subset C of receiving elements. These multiple signals are then combined to generate the ultimate beamformed echo signal.

The Delay-and-Sum (DAS) algorithm serves as the fundamental beamforming algorithm employed in ultrasound imaging. In DAS, each received signal undergoes a time delay adjustment based on the sensor array geometry. Following the time delay application, the signals are assumed to be time-aligned, and the resultant beamformed signal is obtained by summing the time-aligned signals, each weighted equally (with a constant value of 1).

In commercial ultrasound devices aiming for a high frame rate, a low-complexity DAS beamforming algorithm is favored. This approach involves using predetermined static delays for time-of-flight correction of received signals, followed by a summation of the channel data to generate a beamformed signal. However, the simplicity of DAS

comes at a cost—the main lobe width and side lobe width of the resulting beamformed signal are compromised.

Alternatively, more advanced adaptive algorithms like the minimum variance distortionless response (MVDR) exist. With adaptive beamforming, the summation weights dynamically adjust based on the data, yielding improved results compared to DAS. Despite their superior performance [MMP+19], adaptive beamformers like MVDR entail significant computational complexity, making them less suitable for real-time applications.

2.2 Deep Multitask Ultrasound Beamforming

Deep learning techniques have been leveraged to enhance the efficiency of model-based and data-adaptive approaches such as DAS and MVDR, particularly in terms of computational efficiency and image quality improvement. Integrating a data-driven methodology like deep learning can lead to a notable reduction in computational load. For instance, utilizing a neural network for estimating the MVDR output image has demonstrated promising outcomes. In [LCdB+20], researchers showcased the ability to produce images comparable in perceptual quality to MVDR while maintaining computational complexity at $O(n^2)$ as opposed to $O(n^3)$. Furthermore, deep learning enables the consolidation of multiple sequential steps within the image formation pipeline, such as beamforming and denoising, into a single, accelerated neural network.

Within commercial ultrasound devices, various desired output functions are sought. Primarily, there's the ultrasound image itself, followed by denoised and enhanced versions of the original image. Examples include speckle reduction and deconvolution. Traditionally, these post-processing tasks are executed as algorithms operating on the final beamformed images. Commercial ultrasound devices incorporate hundreds or even thousands of such optimizations, making this approach inefficient. Adopting the ultrasound image formation pipeline as a neural network presents an opportunity to estimate the outputs of two-stage classical algorithms within a single neural network. This advancement can significantly boost the framerate. Nevertheless, in commercial ultrasound devices where multiple algorithms are typically integrated, implementing each algorithm as a separate neural network proves inefficient in terms of both storage and computational performance. Moreover, this approach lacks scalability for accommodating a large number of desired outputs. Consequently, a multitask learning approach becomes necessary.

Formally, our objective is to train a unified neural network capable of mapping two inputs—channel data and task number—to task-specific desired outputs.

$$f(x, i; \theta) = \hat{y}_i, \tag{2.1}$$

Where x is the channel data, i is the task index, and $\hat{y}_i$ is the estimated task-specific

beamformed signal.

Within the multitask learning framework, a unified neural network can be designed to generate multiple variations of a given ultrasound image. When implemented correctly, this approach offers significant computational advantages, as image generation is no longer a sequential process—the model learns to map channel data directly to the target image without first reconstructing the B-mode image. Additionally, it enhances storage efficiency by integrating and compiling all post-processing functions into a single, unified model.

2.3 Performance Measures

To evaluate our proposed method and compare it with other participants in the study, we utilized the same test datasets and evaluation metrics as outlined in CUBDL [BHH+20b]. Our evaluation framework examines image quality at both local and global levels, employing metrics suitable for assessing performance in specific regions as well as the overall image.

2.3.1 Local image quality metrics

The challenge organizers chose a predefined region of interest (ROI) within the target images to evaluate local image quality. Those ROIs include lesions and point targets to evaluate denoising, contrast, and resolution. The final scoring of local image quality is a combination of the following metrics:

$$\text{Contrast} = 20 \log_{10} \frac{\mu_1}{\mu_2}, \tag{2.2}$$

$$\text{CNR} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \tag{2.3}$$

$$\text{gCNR} = 1 - \sum_x \min(f_1(x), f_2(x)), \tag{2.4}$$

$$\text{SNR} = \frac{\mu_0}{\sigma_0}, \tag{2.5}$$

where μ_i, σ_i, and f_i represent the mean, standard deviation, and the histogram of ROI-i.

The final scoring of local image quality is derived from a combination of several metrics, each providing different aspects of the quality of the reconstructed image within the ROIs.

1. Contrast (Equation 2.2): - Contrast is a measure of the difference in intensity between two regions in the image. - It's calculated as the ratio of the mean intensity (μ_1, μ_2) of two ROIs, typically representing areas with different properties (e.g., lesion and background). - The logarithmic scale (in decibels) is used to express the contrast, which allows for easier comparison and interpretation of the values.

2. Contrast-to-Noise Ratio (CNR) CNR measures the relative difference in signal intensity between two regions of interest, normalized by the noise level. It's calculated as the difference in mean intensity $(\mu_1 - \mu_2)$ between the ROIs divided by the square root of the sum of their variances $(\sigma_1^2 + \sigma_2^2)$. CNR provides insight into the clarity of the differences between different regions in the image while accounting for the noise present in the image.

3. Generalized Contrast-to-Noise Ratio (gCNR): gCNR is a metric that evaluates the local contrast enhancement within the ROIs. It's calculated based on the histograms $(f_1(x), f_2(x))$ of the intensity values within the ROIs. gCNR measures the degree of suppression of the weaker intensity values in the histogram, indicating the improvement in local contrast.

4. Signal-to-Noise Ratio (SNR) SNR measures the quality of the signal relative to the background noise level. It's calculated as the ratio of the mean signal intensity (μ_0) to the standard deviation of the noise (σ_0). SNR indicates how well the signal of interest stands out from the background noise, which is crucial for image clarity and interpretation.

2.3.2 Global image quality metrics

Global image quality metrics are used to evaluate the global beamformed image quality compared to ground truth in terms of ℓ_1 and ℓ_2 losses, peak signal-to-noise ratio (PSNR) and normalized cross-correlation (ρ), given by:

$$\ell_1 = \frac{1}{N} \sum_{1}^{N} |x_n - y_n|, \tag{2.6}$$

$$\ell_2 = \sqrt{\frac{1}{N} \sum_{1}^{N} (x_n - y_n)^2}, \tag{2.7}$$

$$\text{PSNR} = 20 \log_{10} \frac{\text{DynamicRange}}{\frac{1}{N} \sum_{1}^{N} (x_n - y_n)^2}, \tag{2.8}$$

$$\rho = \frac{\sum_n (x_n - \mu_x)(y_n - \mu_y)}{\sqrt{(\sum_n (x_n - \mu_x)^2)(\sum_n (y_n - \mu_y)^2)}}, \tag{2.9}$$

where x is the envelope of the beamformed reconstruction and y is the corresponding ground-truth envelope.

Image comparisons were conducted using pixels falling within the dynamic range of -40 to 0 dB, relative to the maximum pixel value of the ground truth image. This range was chosen to prevent the exaggeration of minor differences in magnitude and to avoid penalizing networks that fail to reproduce low-magnitude acoustic clutter below -40 dB. Furthermore, both ℓ_1 and ℓ_2 losses were computed for images in their linear form (referred to as ℓ_1 and ℓ_2, respectively) and in their log-compressed form (referred to as $\ell_1 - \log$ and $\ell_2 - \log$, respectively). The linear scale is more sensitive to high-amplitude variations (such as edges), while the log scale is more responsive to low-amplitude variations (such as speckles). In summary, these global image quality metrics offer quantitative insights that are not influenced by the region selection criteria applied to local metrics.

Chapter 3

Deep-Learning-Based Multitask Ultrasound Beamforming

This chapter presents a method to multitask learning with convolutional neural networks. Our method takes inspiration from deep learning normalization techniques like [IS15; BKH16; SK16]. In Section 3.1 we introduce the received RF data preprocessing steps. In Section 3.2 we present the chosen neural network architecture and our core idea, we also show how our proposed method can be adjusted for real-life scenarios. In Section 3.3 we present our results and compare it to other participants of the CUBDL challenge test set. Finally, in Section 3.5 we summarize this chapter.

3.1 Channel Data pre-processing

The received demodulated RF echoes undergo initial conversion into in-phase and quadrature (IQ) signals. These IQ signals are typically represented as complex tensors, where the complex component is derived from the Hilbert transform of the demodulated RF recorded signal. This transformation enables the separation of the RF signal into its in-phase and quadrature components, providing a comprehensive representation that captures both amplitude and phase information essential for further signal processing and analysis.

The IQ signals possess the following dimensional characteristics: IQ $\in \mathbb{C}^{E \times C \times N_t}$. Subsequently, time-of-flight correction is applied to the IQ signals. The delay for each pixel in each transmit event is computed based on the geometry of the transducer. For the case of plane-wave imaging setups, where the transducer elements are uniformly and linearly spaced, the delay can be expressed as:

$$\tau_{c,e,r} = \frac{\|r - r_e\| + \|r - r_c\|}{v_s}, \tag{3.1}$$

where c is the receive channel, e is the transmit channel, and r is the pixel location within the x-z grid. The speed of sound v_s is assumed to be constant.

After applying the appropriate delay to each received element, the resulting tensor is time aligned, $\mathbf{IQ}_{tof} \in \mathbb{C}^{E \times C \times N_x \times N_y}$, Finally the beamforming algorithm is applied on the time of flight corrected channel data $\mathbf{IQ}_{tof}$.

3.2 Multitask fully convolutional architecture

3.2.1 Neural Network Architecture

We've designed a custom fully convolutional neural network (CNN) tailored to the task of ultrasound image formation (Figure 3.1). Our model comprises six convolutional blocks, with each block consisting of two convolutional layers followed by GELU activation [HG20]. The input tensor shape of the network is $2 \times C \times W^2$, where 2 represents the input IQ channel dimension, C denotes the number of receiving channels, and W is the width and height of a patch, flattened to W^2. Our model, as depicted in Figure 3.1, implements a channel compounding function estimator. Each convolutional layer employs a 3×1 kernel, with the horizontal axis of the convolutional kernel sliding through the input tensor's receive channel axis. Given the fully convolutional nature of our model, it enables us to estimate the multi-angle acquisition patch-wise, resulting in significantly improved computational efficiency compared to pixel-wise processing.

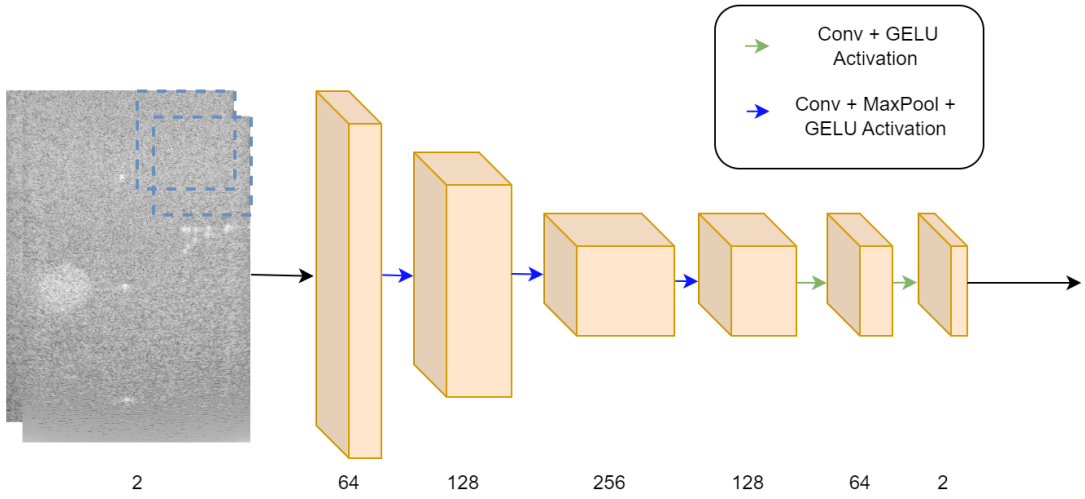

Figure 3.1: The proposed multitask beamforming neural network: pre−processed IQ data are fed to our fully convolutional neural network. The network outputs an IQ estimation corresponding to a task−specific output.

3.2.2 Multitask Learning Approach

The neural network can be modeled as:

$$\hat{\mathrm{I}} = f(X, k; \theta), \tag{3.2}$$

In the given equation, $\hat{I}$ represents the estimated multi-angle reconstruction, while X denotes the pre-processed IQ data, which has been adjusted for factors such as time-of-correction and normalization. The parameter k signifies the task index, while θ stands for the model parameters. Initially, our primary task (where k = 0) involves the reconstruction of multi-angle DAS beamforming images from a single-angle acquisition. Following the completion of training for this foundational task, we proceed to adjust the trained weights for each convolutional layer, denoted as w_i, through the application of a linear scale and bias transformation. This process ensures that the model adapts to variations in input data and task requirements, thereby enhancing its overall performance across different scenarios and tasks. Thus, for each task, the base weights are adapted according to the following transformation:

$$w_i' = \frac{w_i}{s_{i,k}} + b_{i,k}, \tag{3.3}$$

Where $s_{i,k}$ is the scale parameter and $b_{i,k}$ is the bias parameter for layer i and task k, denoted as $s_{i,k}, b_{i,k} \in \mathrm{R^n}$, respectively, where n represents the number of convolutional filters in layer i.

During the training of the model, division and summation are carried out element-wise, specifically at the channel dimension. This approach ensures that adjustments are made individually for each channel within the convolutional layer.

By implementing layer-wise per convolution filter normalization, the learned filters can effectively be readapted to suit new tasks or modified objectives. For instance, this technique can prove invaluable in scenarios involving denoising, image enhancement, or sub-sampling, where precise adjustments at the filter level are crucial for achieving desired outcomes.

3.2.3 Controlling Task Specific Output Intensity

In the equation (3.3), we see that the modified convolution weights for each task are influenced by scale and bias parameters linked to individual convolution kernels. When it comes to tasks like reducing speckle noise and de-convolution, it's useful to control how much the denoising affects the image (See Figure 3.2). To address this, we suggest a new version of our weight normalization scheme.

$$w_i' = \frac{w_i}{s_{i,k}^{\alpha}} + \alpha b_{i,k}. \tag{3.4}$$

Where α is the task intensity control factor, defined within the range $0 \leq \alpha \leq 1$.

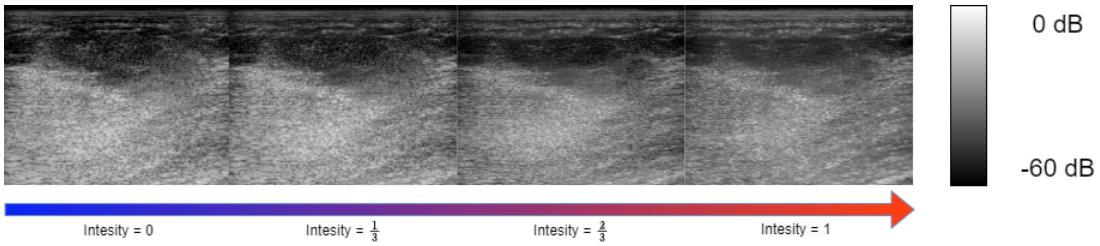

Figure 3.2: Controlling the de−speckling effect. The output is identical to the base task for $\alpha = 0$. The output is a full task-specific effect for $\alpha = 1$. By choosing different α values, we can control the amount of convolution kernel weights scale and bias, and hence control the de−speckling effect.

When $\alpha = 1$, the target task is applied with maximal effect, fully influencing the outcome. Conversely, when $\alpha = 0$, the target task is completely eliminated, exerting no effect on the result. For values of α within this range, the output is an interpolation between the target task and the base task. This means the output effect is proportionally reduced according to the value of α, resulting in a smooth transition between the target task and the base task based on the specified intensity.

3.3 Experimental Setup

3.3.1 Dataset

Our model undergoes training using a dataset detailed in [BHH+20a]. Out of the 28 raw RF samples available for training, 80% are allocated to the training set, while the remaining 20% are reserved for the validation set. This careful partitioning ensures that the model is exposed to substantial data during training while having a dedicated portion for performance assessment and hyperparameter tuning during the validation phase. For model evaluation, we employ the test set from the CUBDL, which encompasses data gathered by multiple institutions. This dataset includes phantom, in vivo, and simulation samples, offering a diverse range of scenarios for robust evaluation. Including data from various institutions and different types of samples helps assess the model's generalizability across different conditions and acquisition setups. All training and test data involve plane wave transmission, utilizing either 31 or 75 angles within an acquisition angle range of $-15°$ to $15°$ or $-16°$ to $16°$. This broad range of acquisition angles ensures the model can handle various imaging geometries, which is crucial for its applicability in real-world scenarios.

The raw RF data samples originate from three distinct scanner models: Verasonics Vantage 128, Verasonics Vantage 256, and ULA-OP 256. Including multiple scanner models introduces variability in the dataset, helping the model learn to generalize across different hardware configurations. Furthermore, six different transducer types contribute to the dataset, ensuring diverse acquisitions to mitigate overfitting. The

variety in transducer types exposes the model to different frequency responses and imaging characteristics, which enhances its robustness.

The center frequencies of the data range from 3.1 to 8 MHz, covering a broad spectrum that includes both low and high-frequency ultrasound imaging. This range allows the model to be trained on data representing different tissue types and imaging depths. The sampling frequencies span from 6.25 to 78.125 MHz, providing high-resolution data to capture fine details in the images. High sampling frequencies are particularly important for applications requiring detailed structural information. A comprehensive description and detailed information about the test and train data are available in [HWG+21b], providing an in-depth understanding of the dataset's characteristics. The detailed dataset description includes information on the acquisition protocols, the specific characteristics of the phantom, in vivo, and simulation samples, and the preprocessing steps applied to the raw RF data.

3.3.2 Target Samples Generation

On each task, our neural network is trained to predict the IQ tensor of the relevant task. We train our multi-task beamformer for three different tasks:

1. Base task: Multi-angle reconstruction from single-angle acquisition. The reconstruction of multi-angle acquisition from a single-angle acquisition yields a notable increase in performance. This is attributed to the faster sensor data acquisition from a single angle compared to a multi-angle acquisition setup. Furthermore, the beamforming algorithm benefits from increased efficiency as the dimensionality of the input signal is reduced. Both the multi-angle and single-acquisition samples are generated using DAS. For the single-angle acquisition samples, the target IQ tensor is beamformed with only a single center angle of 0°.

2. Speckle noise denoising: In most commercial ultrasound devices, there is an image denoising algorithm as post-processing to the traditional beamforming after the conventional steps of envelope detection and log compression. Integrating the image-denoising step into the beamformer leads to further performance improvements. We generated our ground-truth samples for speckle noise reduction by applying [CHKB09], a variation of the Bayesian non-local-means image denoising algorithm, to the multi-angle DAS image formation pipeline output.

3. Sub-sampled channel data reconstruction: Training the beamformer to reconstruct the IQ tensor from subsampled RF data in the channel axis introduces both performance and cost improvements, as it allows the usage of ultrasound probes with fewer transducer elements. Specifically, we applied a deterministic $\times 2$ sub-sampling to our input IQ samples; we zeroed out all the transducer element recordings at even indexes, feeding our model with only sensor reading from odd transducer element indexes.

3.3.3 Implementation Details

1. Base task: First, we train the neural network for our base task of multi angle reconstruction from single angle acquisition; we used AdamW optimizer [LH17], with a learning rate of 3×10^{-4}, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The network is trained for 50 epochs with a step learning rate scheduler reducing the learning rate by a factor of 0.8 every 10 epochs. The batch size was set to 32. Each time-of-flight corrected RF acquisition sample is split into square patches with a width and height of 16 pixels, resulting in a total of 52,889 training samples and 13,070 validation samples. To further increase the data variety, custom data augmentations are used:

   (a) Row flipping: both the output and input IQ patch rows are flipped (mirrored)

   (b) Column flipping: both the output and input IQ patch columns are flipped (mirrored).

   Each augmentation has been applied randomly with a probability of 30%.

2. Sub-tasks: After training for the base task, we chose the best-performing weights on the validation set, and then we trained only the convolution scale and bias parameters on each sub-task. The learning rate was set to 1×10^{-5}, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the network convolution normalization parameters were trained for 20 epochs for each task. All the model training and experiment code was performed with the Pytorch [PGC+17] framework.

For both cases, our loss function was a linear combination of $\log l1$ and MS-SSIM [WSB03] loss, given by:

$$L = \frac{1}{N}\sum_{i}^{N} |\log_{10}(\hat{y}_i) - \log_{10}(y_i)| + \lambda * \text{SSIM}(\hat{y}_i, y_i), \tag{3.5}$$

where $\hat{y}$ is the model envelope estimation, y is the ground-truth, and SSIM is the MS-SSIM loss.

The $\log l1$ component of the loss is used to ensure an accurate estimation of the envelope of the signal, and the MS-SSIM component ensures that the statistical and structural properties of the image such as the contrast and luminance remains close the the ground-truth. λ is set at 0.1.

3.4 Experimental Results

3.4.1 Ablation Study

In their research, Goudarzi *et. al.* [GAR20] introduced a neural network architecture based on MobileNetV2 [**inproceedings**] for reconstructing a multi-angle DAS beam-

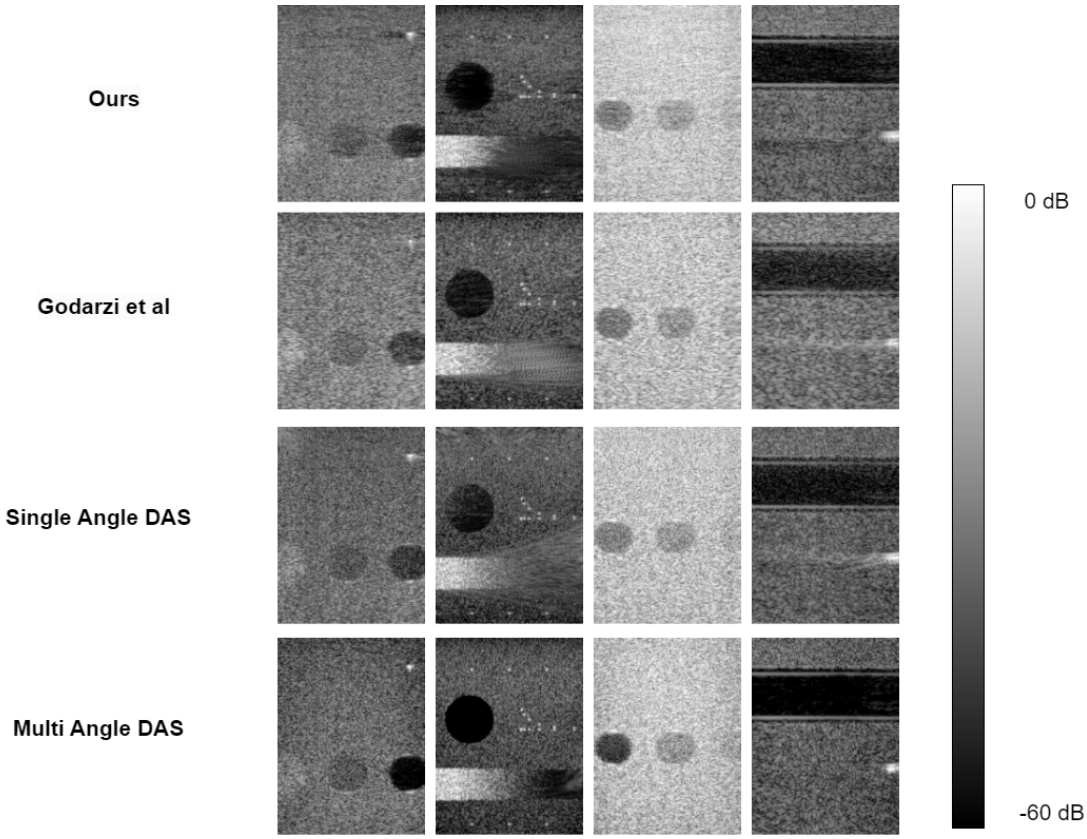

Figure 3.3: Test set samples of our base task multi-angle reconstruction from single-angle acquisition. Our model can remove most of the noise and scattering $\log \text{SpeckleSNR}$ of 2.299 and ρ of 0.93 − outperforming all the other challenge participants (Tables 3.1 and 3.2)

former from single-angle acquisitions. The network operates on a $2 \times C \times W$ tensor, where C denotes the number of receiving channels, with W fixed at 32 for the spatial window size. Its output is a concise two-element vector representing the IQ components of the estimated multi-angle DAS. With 2.226 million parameters, MobileNetV2 is recognized for its lightweight design, facilitating faster computation and inference. However, its pixel-by-pixel reconstruction method poses limitations for real-time applications due to its processing speed and efficiency challenges.

Rothlübbers *et. al.* [RSE+20] took a distinct approach by leveraging the neural network's output as beamforming weights instead of directly estimating multi-angle IQ components. These weights are applied to the single-angle input to generate the multi-angle estimate. Their training dataset comprises 107 samples of privately acquired raw ultrasound RF data along with publicly available data [LRC+16]. The network is trained using a combination of mean squared error per pixel loss and multi-scale structural similarity (MS-SSIM) [WSB03].

After performing beamforming and log compression of the received echo signal, the subsequent step in the ultrasound image pipeline involves post-processing. This crucial phase aims to enhance contrast and reduce noise in the beamformed signal. Noise

reduction is particularly essential in scenarios like plane wave ultrasound transmission, where inherent higher noise levels and lower spatial resolution are commonplace compared to focused transmit ultrasound. In medical imaging, these post-processing techniques automate diagnostics and elevate image quality. Denoised images offer clearer visualizations of anatomical structures and pathological features, thereby enhancing the precision and efficiency of medical diagnoses.

Bhatt *et. al.* [BNKL20] proposed a UNet-based architecture [RFB15] designed to predict both segmentation and image reconstruction. Their model architecture includes a single encoder and two decoders, each dedicated to producing task-specific outputs. The first decoder reconstructs ultrasound images, while the second generates segmentation maps. A significant advantage of their approach lies in simultaneously producing both types of outputs. However, the model's computational requirements escalate with additional tasks due to the necessity of training separate encoders from scratch for each task.

Khan *et. al.* [KHY21b] presented an alternative strategy employing a U-Net variation augmented with adaptive instance normalization layers (AdaIN) [HB17] at the bottleneck block. In addition to the primary U-Net beamformer, they trained a compact, fully connected neural network to adjust AdaIN parameters (normalization mean and variance) tailored to specific task demands. This method enhances scalability and performance by selectively modifying the bottleneck layer representation for each task, effectively controlling task-specific outputs.

In summary, while each approach offers unique advantages in terms of computational efficiency, image quality enhancement, and task-specific adaptability, they also present challenges such as increased computational demands with task complexity and limitations in real-time processing capabilities. These considerations underscore the ongoing evolution and optimization efforts in ultrasound image reconstruction and post-processing methodologies.

3.4.2 Image Targets

This section evaluates our model qualitative performance using the test set from [HWG+21a]. We also compare our result to other participants of the CUBDL [HWG+21a] challenge [GAR20; RSE+20]. We measure global and local image quality metrics described in Section 2.3.

Our model's performance, as evaluated against existing methodologies using data from Tables 3.1 and 3.2, reveals compelling advancements in image quality metrics. These findings not only highlight our model's strengths but also shed light on its implications for practical applications in medical imaging.

Model	Contrast dB	Speckle CNR	gCNR	log SNRS	x fwhm	z fwhm
Goudarzi [GAR20]	−13.77	1.345	0.814	1.827	**0.0003**	0.0004
Single Plane	−10.5	1.05	0.674	1.85	0.0004	0.0004
Ours	**−15.84**	**1.632**	**0.88**	**2.299**	0.0005	**0.0003**
Ours— ×2 Sub Sampled	−5.07	1.083	0.62	1.94	0.0006	0.0003
Single Plane— ×2 Sub Sampled	−4.95	0.542	0.394	1.842	0.0004	0.0004
Ground-Truth— Multi-Plane	−24.70	1.53	0.946	1.806	0.0004	0.0003

Table 3.1: CUBDL [HWG+21a] test set local results. Local image quality results from the CUBDL test set. Our model outperforms the competitors in the challenge by a significant margin in every measure except spatial resolution along the x-axis (x fwhm)

3.4.3 Local Image Quality Metrics

Table 3.1 outlines several key metrics crucial for evaluating local image quality. Our model, denoted as "Ours," consistently outperforms competitors such as Goudarzi [GAR20] and a single-plane approach across various parameters. Notably, our model achieves superior Contrast, Speckle CNR, gCNR, and log SNRS compared to other models in the CUBDL test set. These results signify enhanced clarity and reduced noise, which are critical for precise medical imaging diagnoses.

The implications of these results are profound. By achieving higher contrast dB and speckle CNR, our model enhances the visibility of anatomical structures in ultrasound images. This improvement is particularly crucial in scenarios where image clarity directly impacts diagnostic accuracy. Additionally, the superior gCNR and log SNRS indicate better overall image quality, which can lead to more reliable diagnostic interpretations and fewer misdiagnoses.

3.4.4 Global Image Quality Metrics

Table 3.2 presents global image quality metrics, focusing on ℓ_1 and ℓ_2 norms, logarithmic norms ($\log \ell_1$ and $\log \ell_2$), PSNR, and correlation coefficient (ρ). Our model consistently outperforms Goudarzi [GAR20] and the single-plane approach across these metrics,

Model	ℓ_1	$\log \ell_1$	ℓ_2	$\log \ell_2$	PSNR	ρ
Goudarzi [GAR20]	0.03	0.42	0.05	0.56	29.10	0.91
Single Plane	0.03	0.39	0.042	0.53	30.36	0.93
Ours	**0.029**	**0.36**	**0.0408**	**0.49**	**30.4**	**0.93**
Ours—$\times$2 Sub Sampled	0.035	0.41	0.0504	0.55	28.52	0.9
Single Plane—$\times$2 Sub Sampled	0.0433	0.525	0.059	0.7	27.43	0.86

Table 3.2: CUBDL [HWG+21a] test set global results. Global image quality results on multi-angle DAS reconstruction from single-angle acquisition, and image reconstruction from 2x sub-sampled channel data. our model produces the best results in every measurement. Additionally, for the task of image reconstruction from sub-sampled channel data, our model outperforms DAS algorithm.

showcasing its ability to minimize reconstruction errors and maximize signal fidelity.

The implications of these findings extend to practical applications in medical imaging. By reducing ℓ_1 and ℓ_2 norms, our model enhances reconstruction accuracy, resulting in clearer and more detailed images essential for medical professionals. Moreover, the higher PSNR values indicate improved image fidelity, which is critical for identifying subtle anatomical features and abnormalities.

3.4.5 Practical Considerations

Since our model is fully convolutional neural network in each forward pass a patch of pixel is reconstructed. This approach led to a substantial computational performance increase since, in each forward pass, our model generated 256 pixels instead of only one. Our approach of estimating the beamformed IQ samples in patches instead of single pixels at a single forward pass yields a fast inference. Since in each forward pass we process W^2 more pixels at a single forward pass where W is the window length.

3.4.6 Multi-Task Results

We evaluate our multi-task approach by adapting our trained model to two different tasks: transducer elements sub-sampling and speckle noise reduction (despackling). We employed the same evaluation metrics as in the base task test section 3.4.2.

3.4.7 Sub-Sample reconstruction results

In our methodology, a crucial step involves sub-sampling the input IQ cube at the channel dimension, employing a downsampling rate of $\times 2$. This process entails selectively retaining only the elements at odd indexes within the IQ cube, while all other elements are systematically removed, effectively setting them to zero. By implementing this sub-sampling strategy, we aim to reduce the computational complexity and memory requirements of subsequent processing stages, facilitating more efficient data handling without significantly compromising the overall reconstruction quality.

It's noteworthy that our choice of retaining elements at odd indexes serves a specific purpose, aligning with established practices in signal processing and image reconstruction. This selective retention allows us to maintain essential information while still achieving a significant reduction in data size. Furthermore, this approach is consistent with the principles of sub-sampling, where the objective is to strike a balance between data reduction and preservation of pertinent information.

Following the sub-sampling process, we conducted a comprehensive evaluation of our reconstruction model's performance, leveraging the same testing framework utilized in the previous section 3.4.2. Our assessment focused on both local and global image quality metrics, providing insights into how our sub-sampled reconstruction model compares with traditional single-angle DAS methods.

From our analysis, as illustrated in Table 3.2, it becomes evident that our sub-sampled reconstruction model achieves competitive global image target results when compared to prior studies such as [GAR20]. Moreover, our model exhibits a notable improvement over single-angle sub-sampled reconstruction methods, showcasing its efficiency in mitigating the performance degradation associated with sub-sampling.

Delving deeper into the local image quality assessment, as depicted in Table 3.1, our model maintains a superior performance compared to single-angle DAS reconstruction, even after sub-sampling. However, it is worth noting that our model's performance in terms of contrast falls short when compared to fully sampled reconstructions, highlighting an area for potential refinement and optimization in future iterations.

3.4.8 Speckle Noise Reduction

Our evaluation of speckle noise reduction involved testing both global and local image quality metrics, as described in (2.2) through (2.9). In our assessment, we compared the performance of our neural network-based speckle-reduced images against the speckle denoising algorithm proposed in [CHKB09], applied to both multi-angle (considered as ground-truth) and single-angle reconstruction scenarios.

The results of our global image quality test, as presented in Table 3.4, underscore the superiority of our model over the speckle-denoised single-angle DAS-generated images in terms of various global image quality metrics. Notably, our model showcases a significant improvement, highlighting its effectiveness in mitigating speckle noise artifacts

Model	Contrast dB	Speckle CNR	gCNR	Speckle SNR
Single-Angle + speckle reduction	−10.5	1.05	0.674	1.85
Ours: speckle reduction	−10.07	1.673	0.885	**3.17**
Multi-Angle + speckle reduction (ground-truth)	**−25.6**	**2.423**	**0.98**	2.713

Table 3.3: CUBDL [HWG+21a] test set local result on speckle reduction task. Our multitask learning approach has been compared to single-angle and multi-angle DAS with post-processing. The model is trained to estimate the output of multi-angle DAS from single-angle acquisition input. We notice that the ground truth of multi-angle DAS with speckle reduction still yields the best results, however, our approach suppresses single-angle DAS with post-processing by a significant margin.

and the effectiveness of out multitask learning method.

Upon closer examination of the local image quality metrics, detailed in Table 3.3, it becomes evident that our model consistently outperforms the speckle-reduced single-angle DAS approach across all measurements, with the exception of contrast. While our reconstruction exhibits superior performance in metrics such as gCNR and Speckle SNR, it falls short in contrast to the ground-truth multi-angle DAS with speckle reduction.

Despite the observed inferiority in contrast relative to the ground-truth multi-angle DAS, our reconstruction remains competitive or even superior in other metrics, underscoring its overall effectiveness in reducing speckle noise while preserving essential image details. This nuanced evaluation provides valuable insights into the strengths and limitations of our speckle noise reduction approach, paving the way for further refinement and optimization.

3.4.9 Comparative Analysis and Limitations

For the fundamental task of single-angle reconstruction from multi-angle acquisition, our research has substantiated that our model outperforms the CUBDL challenge competition winner regarding global image quality metrics. Specifically, our model shows state-of-the-art results with all the global image quality metrics, with the local image quality metrics, our model produces results that out-perform CUBDL challenge winner by a noticeable gap, In every measure except for spatial resolution along the x-axis (x-full with at half maximum).

Model	$\log \ell_1$	$\log \ell_2$	PSNR	ρ
Single Plane + speckle reduction	0.57	1.07	24.8	0.593
Ours: speckle reduction	**0.3**	**0.41**	**29.01**	**0.92**

Table 3.4: CUBDL [HWG+21a] test set global result on speckle reduction task. Our multitask learning suppresses single angle DAS with speckle reduction post-processing by a significant margin on every measure.

The network proposed by Goudarzi *et. al.* [GAR20] learns to map between a window of IQ from a single angle acquisition to a single pixel in the center of the window. They trained the model to estimate the corresponding center pixel in multi-angle acquisition. Thus in each forward pass, the output is a single pixel. Our model is trained to learn a mapping between a patch of single-angle acquisition IQ data to a patch of multi-angle beamformed IQ data. Thus, since every layer applies convolution in the channel domain, our model predictions are less focused on a single pixel than the MobileNetV2 beamformer.

The MobileNetV2 model can be parallelized in a practical implementation, thus reconstructing multiple pixels with a single parallelized forward pass. Due to memory limitations, parallelizing all the pixels within an ultrasound scan is not feasible with current technology. Since our neural network reconstructs a patch of pixels within a single forward pass, parallelization of our model to reconstruct the entire image is more feasible.

Unlike conventional image processing tasks that are typically executed by 2D convolutional neural networks, beamforming aims to compound multiple channels of the received signal. Therefore, implementing this objective involves utilizing a one-dimensional convolutional kernel that slides exclusively through the channel dimension. 1d convolution kernels are also faster in terms of computational resources since fewer multiplication and sum operations are performed. Also, they are lighter in terms of parameter count. We then fitted the mean and variance parameters of the layer weight normalization to two different tasks: sub sampling and speckle noise reduction. Our model outperforms sub-sampled single and delay and sum reconstruction in every global and local image quality metric measurement. The sample images in Fig. 3.4 show our sub-sampled reconstruction model ability to remove scattering, resulting from noisy measurement compared to sub-sampled delay and sum reconstruction. This property of our beamformer is also observable with the global image quality metrics, where our noise measurements, such as PSNR and , are significantly better than the sub-sampled DAS reconstruction.

For the next task, we compared our model against single angle delay and sum with

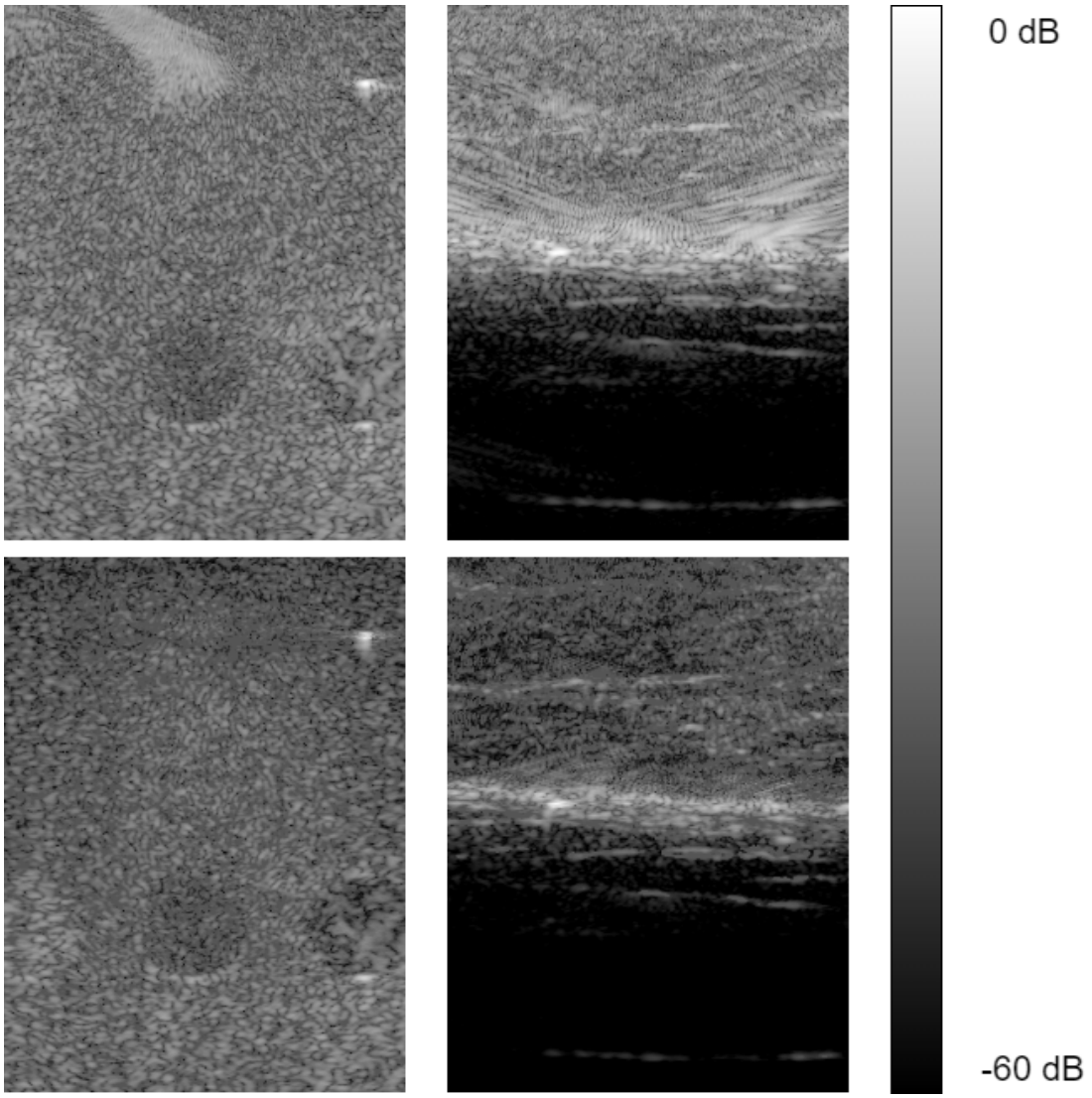

Figure 3.4: The image reconstruction samples of sub-sampled data at the channel dimension are presented. The first row shows the DAS reconstruction from sub-sampled channel data, while the bottom row displays our model reconstruction. Our model demonstrates a reduction in noise in the final image, resulting from the fewer elements used. Additionally, it produces an image with higher contrast compared to the sub-sampled single-angle reconstruction. Both images are samples from CUBDL [BHH$^+$20b] test set.

a speckle noise reduction post-processing. From Table 3.3 we can conclude that our model produces the best results in local image quality metrics compared to single-plane wave reconstruction with speckle denoising post-processing - except for a small gap in contrast. with the global image quality metrics our model outperforms the single-angle DAS in every measurement by a significant amount.

Our model is trained using the CUBDL [BHH$^+$20b] dataset , which consists of ultrasound channel data from various ultrasound machines and probe configurations. These machines and configurations have different center frequencies and sampling frequencies.

Additionally, the dataset includes simulated data, phantom scans, and in-vivo data. This diverse dataset proves the robustness of our model across various experimental settings for plane wave ultrasound image formation. However, since we narrowed our work to plane wave ultrasound our model is limited to only plane wave ultrasound beamforming and is not suitable for other image formation methods like focused transmit.

3.5 Summary

We have presented a method for multitask learning applied to ultrasound beamforming. We started by training our model to reconstruct images from multiple transmission angles from only single center transmission. Then, we applied our method for multitask learning to the learned convolutional filters from our base task, and adapted them to two different tasks: speckle noise reductions, and image reconstruction from sub-sampled data at the receive channel dimension. We then tested our method with the test set of CUBDL challenge. Our model outperforms sub-sampled single and delay and sum reconstruction in every global and local image quality metric measurement. Notably, the global image results of image reconstruction from sub-sampled signal shows noise measurements, such as PSNR and ρ, that are significantly better than the sub-sampled DAS reconstruction.

For the next task, we compared our model against single-angle delay and sum with a speckle noise reduction post-processing. From Table local results Table 3.3 we can conclude that our model produces the best results in local image quality metrics compared to single-plane wave reconstruction with speckle denoising post-processing, except for a small gap in contrast. With the global image quality metrics, our model out performs the single-angle DAS in every measurement by a significant amount.

Chapter 4

Lightweight Multitask Deep Learning for Ultrasound Image Formation

This chapter presents a method to efficient multitask learning. In designing our multitask approach, we prioritize minimizing the number of learnable parameters and enhancing computational efficiency during inference. This ensures its practical applicability in real-life scenarios. In Section 4.1 we introduce ultrasound image formation and our proposed improved method for multitask learning. In Section 4.2 we present the experimental setup and the results of the CUBDL challenge. In Section 4.3 we perform a theoretical analysis of the parameter overhead required for our method. Finally, in Section 4.4 we summarize this chapter.

4.1 Multitask learning with linear filter Transformation

In this work, we apply our proposed method to plane wave ultrasound transmission [CFT12]. In plane-wave ultrasound, all the ultrasound transducer elements are used both for transmission and receive. The sum of all the waves from the different transducer elements forms a plane wave. Each transmission event E is then a different angle of the plane wave, Where each angle is formed by applying the requisite delay to each transducer element excitation signal. This study introduces an innovative and efficient approach to multitask learning within the domain of ultrasound beamforming. Initially, a neural network is trained for the fundamental task of ultrasound image reconstruction. More precisely, the neural network is trained to predict the output of the DAS algorithm using raw RF samples. The training dataset consists of 128 channels and 75 transmit events, focusing specifically on a single transmit event from the center angle.

After the initial training for ultrasound image reconstruction, the model is adapted for the additional task of speckle noise reduction. Since the proposed architecture

includes convolutional filters that operate on the data's channel dimension, this model implements a non-linear weighted channel summation model. The non-linearity of the model allows for the estimation of various task variations with minor changes to the internal statistics of the data. As demonstrated in StyleGAN [KLA19], per-instance modification of internal data statistics using adaptive instance normalization can lead to significantly different outputs. A similar approach has been shown in [KHY21a] in the context of multitask ultrasound beamforming.

This work proposes an alternative method to alter the data representation. While adaptive instance normalization is limited to scaling and centering the data across new means and variances, our approach adapts the convolutional filters of the entire model. This method allows us to comprehensively change the data representation, enabling the model to learn a more robust representation. Adapting the convolutional filters gives the model a more nuanced and robust feature adaptation capacity. This enhanced adaptability can improve performance and better accommodate new tasks, making our method a promising alternative to traditional adaptive normalization techniques.

Formally, the learned filter weights for the Ith convolutional layer are denoted as W_i where $W_i \in \mathbf{R}^{C_1 \times C_2 \times k_1 \times k_2}$, where C_1 and C_2 represent the input and output channels of the layer, respectively, and K_1 and K_2 represent the dimensions of the kernel. Then, for each layer, we learn (optimize, with stochastic gradient descent) an optimal linear transformation matrix $T \in \mathbf{R}^{(K_1 \cdot K_2 \times K_1 \cdot K_2)}$ such that

$$W_{ij} = W_i \times T_{ij}, \tag{4.1}$$

where W_{ij} is the transformed matrix at layer i for task j and T_{ij} is transformation matrix for layer i and task j.

4.1.1 Model Architecture

Layer type	input shape	output shape	Activation
2D Conv Layer	$2 \times Nc \times 256$	$64 \times Nc \times 256$	GELU
2D Conv Layer	$64 \times Nc \times 256$	$64 \times Nc \times 256$	GELU
Max pooling	$64 \times Nc \times \times 256$	$64 \times \frac{Nc}{2} \times 256$	-
2D Conv Layer	$64 \times \frac{Nc}{2} \times 256$	$128 \times \frac{Nc}{2} \times 256$	GELU
2D Conv Layer	$128 \times \frac{Nc}{2} \times 256$	$128 \times \frac{Nc}{2} \times 256$	GELU
max pooling	$128 \times \frac{Nc}{2} \times 256$	$128 \times \frac{Nc}{4} \times 256$	-
2D Conv Layer	$128 \times \frac{Nc}{4} \times 256$	$256 \times \frac{Nc}{4} \times 256$	GELU
2D Conv Layer	$256 \times \frac{Nc}{4} \times 256$	$128 \times \frac{Nc}{2} \times 256$	GELU
max pooling	$256 \times \frac{Nc}{4} \times 256$	$256 \times \frac{Nc}{8} \times 256$	-
2D Conv Layer	$256 \times \frac{Nc}{8} \times 256$	$128 \times \frac{Nc}{8} \times 256$	GELU
2D Conv Layer	$128 \times \frac{Nc}{8} \times 256$	$128 \times \frac{Nc}{8} \times 256$	GELU
2D Conv Layer	$128 \times \frac{Nc}{8} \times 256$	$64 \times \frac{Nc}{8} \times 256$	GELU
2D Conv Layer	$64 \times \frac{Nc}{8} \times 256$	$64 \times \frac{Nc}{8} \times 256$	GELU
2D Conv Layer	$64 \times \frac{Nc}{8} \times 256$	$2 \times \frac{Nc}{8} \times 256$	Linear

Table 4.1: Fully Convolutional Beamformer Architecture.

A specialized fully convolutional neural network (CNN) [LLY+21], Table 4.1, is utilized to address the challenges in ultrasound image formation. Our model comprises six convolutional blocks, with each block consisting of two convolutional layers, followed by GELU activation [HG16]. The input tensor shape of the network is $2 \times C \times W^2$, where 2 represents the input IQ dimension, C is the count of receiving channels, and W^2 is a patch with both width and height equal to W flattened. Our model implements a channel compounding function estimator.

For each convolutional layer, we utilize a 3×1 kernel, where the horizontal axis of the convolutional kernel slides through the input tensors' receive channel axis. The fully convolutional nature of our model enables the patch-wise estimation of multi-angle acquisition, resulting in significantly faster processing compared to pixel-wise approaches. The complete inference pipeline is shown in Fig. 4.1.

4.2 Experimental Results

To evaluate the effectiveness of our proposed method, we conducted a comprehensive experimentation process. Initially, we trained two models, both with identical architectures. The first model underwent training for the base task of ultrasound beamforming,

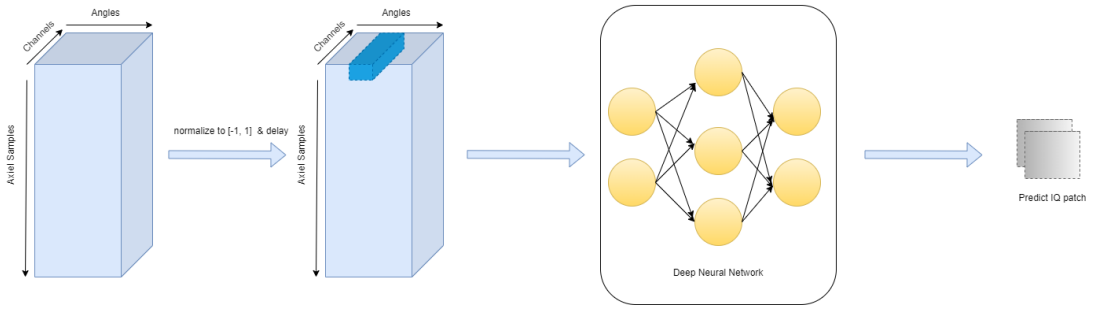

Figure 4.1: Proposed ultrasound image formation pipeline. The input raw channel data is pre-processed with time of flight correction, and z-score normalization. Then, IQ patches are fed into the neural network to perform target task estimation.

followed by further optimization using our proposed method tailored for speckle reduction. Subsequently, the second model was specifically trained for speckle reduction, using the same architecture and dataset. However, in this case, all model weights were made trainable, enabling a more specialized learning process.

To rigorously assess the performance of our approach, we used the test set provided in [BHH+20a], which includes benchmarks for both local and global image quality evaluations. To evaluate the effectiveness of our speckle reduction technique, we applied the algorithm described in [CHKB09] to the test set images sourced from [BHH+20a].

For a comprehensive analysis of our approach's results, we compared it against several baseline methods. These include training the same model architecture directly on speckle reduction data, a model trained using our proposed multitask approach, and single-angle as well as multi-angle DAS image formation with post-processing of speckle denoising algorithm, detailed in [CHKB09]. The results obtained from each approach, covering both local and global image quality metrics, are succinctly summarized in Tables 4.2 and 4.3.

4.2.1 Global Image Quality Metrics

The global image quality metrics for the different approaches are summarized in Table 4.2. These metrics include ℓ_1, $\log \ell_1$, ℓ_2, $\log \ell_2$, PSNR, and the Pearson correlation coefficient (ρ).

The results indicate that the Base model achieved the best performance across several metrics. Specifically, it recorded the lowest ℓ_1 (0.03) and ℓ_2 (0.043) errors, which indicate more minor deviations from the ground truth images. It also achieved the highest PSNR (30.4 dB), which implies a higher fidelity in the reconstruction, and the normalized cross-correlation ($\rho = 0.94$), suggesting a strong linear relationship between the reconstructed and ground truth images.

The multitask model showed slightly higher errors ($\ell_1 = 0.03$, $\ell_2 = 0.048$) and marginally lower PSNR (29.5 dB) and ρ (0.93) compared to the base model. However, it still performed significantly better than the single-angle DAS with speckle denoising,

Model	ℓ_1	**log** ℓ_1	ℓ_2	**log** ℓ_2	**PSNR**	ρ
Base	0.03	0.275	0.043	0.37	30.4	0.94
Multitask model	0.03	0.3	0.048	0.41	29.5	0.93
Single Angle DAS + Speckle Denoising	0.043	0.57	0.1	1.07	24.78	0.593

Table 4.2: Speckle Reduction: Global Image Results.

Model	**Contrast**	**CNR**	**gCNR**	**Speckle SNR**
Base	-14.05	-1.95	0.89	3.14
Multitask model	-13.32	-1.74	0.89	2.71
Single Angle DAS + Speckle Denoising	-10.5	1.05	0.674	1.85
Ground Truth	-25.6	2.423	0.98	2.713

Table 4.3: Speckle Reduction: Local Image Results.

demonstrating that our multitask learning approach effectively maintains high global image quality while being adaptable to the target task.

The single angle DAS + speckle denoising algorithm as post-processing, had the highest errors ($\ell_1 = 0.043$, $\ell_2 = 0.1$) and the lowest PSNR (24.78 dB) and ρ (0.593), indicating it was the least effective in preserving global image quality. This underscores the advantage of our proposed multitask learning approach over traditional speckle reduction methods.

The multitask model's results closely align with the base model, suggesting that our proposed multitask learning approach effectively preserves global image quality while being more adaptable and potentially more scalable.

4.2.2 Local Image Quality Metrics

Local image quality metrics, which include Contrast, CNR, gCNR, and speckle SNR, are detailed in Table 4.3. The findings are as follows: The base model achieved the highest performance in several metrics. It had the best CNR (-1.95) and speckle SNR (3.14), indicating superior contrast differentiation and speckle noise reduction. The multitask model showed slightly reduced performance compared to the base model, with CNR (-1.74) and speckle SNR (2.71) marginally lower. However, it performed better than the single-angle DAS with speckle denoising and showed comparable performance in gCNR (0.89).

Single-angle DAS with speckle denoising demonstrated the lowest values in Contrast (-10.5), CNR (1.05), and speckle SNR (1.85), indicating the least effective speckle reduction and image quality. The ground truth sample, which refers to multi-angle

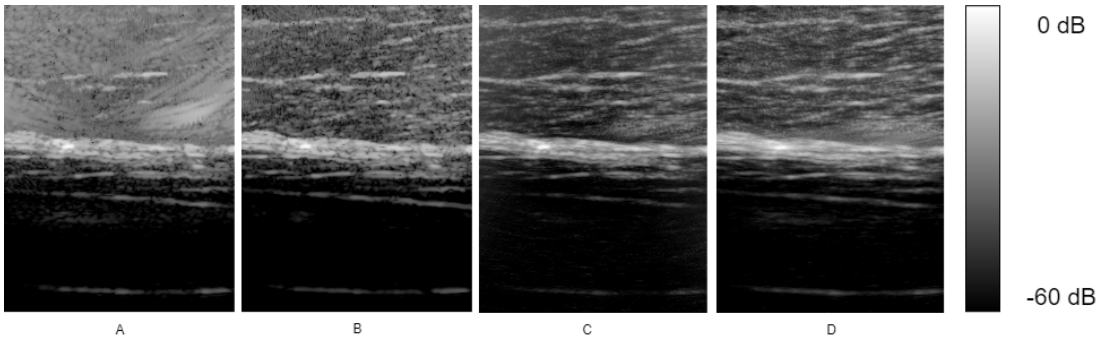

Figure 4.2: In vivo inference of the different approach of speckle reduction. A) Single angle DAS with speckle reduction. B) 75-angle DAS with speckle reduction. C) Specifically trained neural network for speckle reduction. D) Our multitask learning approach of speckle reduction. The results are very close to those of the specifically trained model while adapting the model to the target task with significantly fewer parameters, thus providing a more scalable approach.

DAS and speckle denoising included for reference, shows significantly different metrics highlighting the performance gap that any algorithm must bridge to achieve ideal results. The multitask model's local image quality metrics are notably superior to the single-angle DAS approach and closely approximate those of the base model. This supports the effectiveness of our multitask approach in enhancing local image quality while maintaining the benefits of a more generalized learning process.

Figure 4.2 illustrates in vivo inference results for different speckle reduction approaches. The visual comparison shows: Single-angle DAS with speckle denoising images produced by this approach exhibits more speckle noise and less clear structural details, reflecting its lower performance in global and local image quality metrics. Multi-angle DAS with speckle denoising demonstrates the highest quality. The contrast between the different regions of the scanned tissue appears to be the best. This algorithm is used as ground truth for our experiments. Specifically, the trained model for speckled denoising provides high-quality images with effective speckle noise reduction, serving as a benchmark for evaluating the multitask approach. Lastly, the multitask learning approach delivers visual results very close to those of the specifically trained neural network, confirming its effectiveness in reducing speckle noise while preserving important image details. This also highlights the efficiency of our approach, which adapts to the target task with significantly fewer parameters, making it a more scalable solution.

Analyzing global and local image quality metrics demonstrates that our proposed multitask learning approach for speckle reduction in ultrasound imaging effectively balances performance and scalability. The multitask model closely matches the performance of a model explicitly trained for speckle reduction and significantly outperforms the single-angle DAS with a speckle denoising approach. These results underscore the potential of multitask learning to enhance image quality in ultrasound imaging ap-

plications while maintaining a more efficient and scalable model architecture. Our proposed multitask learning approach proves to be a promising strategy for speckle reduction, achieving competitive performance in both global and local image quality metrics compared to models trained explicitly for this task. The ability to adapt with fewer parameters highlights the approach's versatility and suggests its potential for broader applications in medical imaging, where maintaining image quality and computational efficiency are paramount.

4.3 Parameter Efficiency

In our approach, the number of parameters required to transfer a given 2D CNN for different tasks is determined by

$$P = \sum_{i=1}^{N} (K_{xi} K_{yi})^2, \tag{4.2}$$

where K_{xi} and K_{yi} represent the x and y dimensions of the convolutional filter at the i-th layer. Typically, these dimensions fall within the range $1 < K_i < 11$, which ensures that the total parameter overhead remains relatively low when compared to other methods.

To provide a concrete example, consider our proposed architecture, which consists of 250,000 parameters. When adapting this model to the specific task of speckle reduction, our method required only an additional 173 parameters. This addition represents merely 0.07% of the model's total parameter count. Such a minimal increase in total parameter count highlights the efficiency of our approach in terms of parameter overhead.

To further illustrate, let us consider the general case of a 2D convolutional layer. The ratio of the number of parameters required by our proposed linear filter weights transformation method to those needed for the convolutional filter weights can be expressed as:

$$R = \frac{K_x K_y}{C_1 C_2 + \frac{C_2}{K_x K_y}}. \tag{4.3}$$

In this formula, C_1 denotes the input channel dimension, C_2 denotes the output channel dimension (number of filters), and K_x and K_y represent the spatial dimensions of the filter. In standard CNN design, it is generally expected that $C_1 C_2 \gg K_x K_y$. This inequality ensures that the parameter overhead required for adapting to a new task is very low, making our approach highly efficient. By employing this method, we significantly reduce the number of additional parameters needed for task adaptation. This efficiency is particularly beneficial in practical applications with constrained computational resources and memory. Our method ensures that the overall model complexity remains manageable while achieving high performance across various tasks.

4.4 Summary

We have presented a novel multitask learning approach, specifically tailored for ultrasound beamforming applications. Our method demonstrates the capacity to transfer knowledge effectively across different tasks within ultrasound beamforming, such as image formation and speckle noise reduction. Through qualitative evaluation, our approach exhibits performance comparable to dedicated network training, while significantly outperforming the single-angle DAS method with speckle noise reduction. We have demonstrated the high efficiency of our method in terms of learnable parameters overhead. By combining this parameter efficiency with the perceptual performance showcased in our tests, our method proves to be well-suited for practical deployment in commercial ultrasound machines.

Chapter 5

Conclusions

5.1 Summary

Ultrasound imaging, a crucial diagnostic tool in medical settings, heavily relies on efficient algorithms for image formation. Traditionally, the algorithms employed have been classical and efficient, catering to the demands of real-time imaging. However, as ultrasound technology advances, there is an increasing need for higher frame rates and improved resolution and contrast, pushing the boundaries of computational efficiency. This demand for faster reconstruction times restricts the use of complex algorithms with high computation costs, posing a challenge for traditional ultrasound image formation methods.

In recent years, the field of deep learning has witnessed remarkable progress, demonstrating state-of-the-art results in various domains, including image and signal processing tasks. Recognizing this trend, there has been a growing interest in harnessing the power of deep learning for ultrasound image formation. Moreover, commercial ultrasound devices commonly incorporate post-processing functions such as speckle noise reduction and deconvolution to enhance the quality of captured images.

Addressing the evolving landscape of ultrasound imaging, this research proposes a novel approach to tackle the reduced frame rate issue and enhance beamforming performance using deep learning methodologies. Specifically, the focus is on developing efficient, scalable, and robust multitask learning approaches tailored to the challenges encountered during the beamforming stage of the ultrasound image formation process. A noteworthy aspect of this research is its emphasis on scalability and efficiency, which has not been a primary focus in existing literature within the field. By prioritizing these aspects, the proposed method aims to bridge the gap between the demanding computational requirements of modern ultrasound imaging and the capabilities of traditional algorithms. In summary, this research represents a significant step towards advancing ultrasound image formation techniques by leveraging the capabilities of deep learning while addressing crucial challenges related to scalability and efficiency. It not only contributes to the growing body of knowledge in medical imaging but also holds promise for

enhancing the diagnostic capabilities of ultrasound technology in clinical practice. In Chapter 3, we introduced a novel multitask learning scheme tailored for convolutional neural networks (CNNs). Central to our method is a weight normalization scheme designed to enhance the adaptability and effectiveness of the CNN model.

Motivated by the observed impact of altering the statistics (mean and variance) of deep features on model output, as demonstrated in style transfer and semantic image-to-image mapping works, we proposed a convolutional filter normalization approach. Drawing inspiration from instance normalization, we aimed to manipulate the weighting of channel data aggregation in the context of ultrasound image formation, specifically within the beamforming process. We demonstrated that by modifying the statistics of convolutional filters, we can influence the output of the beamformer for various tasks. Additionally, we developed a method to control the intensity of task adaptation effects using a scalar parameter within the range of 0 to 1. This parameter allows users to fine-tune the desired level of denoising or other effects on the base image. To evaluate the effectiveness of our proposed method, we conducted benchmarking using the CUBDL challenge test set. Our results were compared to those of the challenge winners, showcasing the superiority of our base method for subsample ultrasound image formation across all metrics. Furthermore, we conducted experiments to assess the performance of our multitask learning approach compared to single-angle DAS on tasks such as speckle noise reduction and subsampled reconstruction. Our findings underscored the superior performance of our multitask learning approach across these tasks.

Chapter 4 extends the work presented in Chapter 3, focusing on two key aspects: perceptual performance and parameter/computational efficiency. To address these aspects, we introduced a more efficient and lightweight linear filter transformation. This transformation significantly reduces parameter overhead, enhancing computational efficiency with improvement in the numerical and perceptual quality measures of the generated images. Through testing on the same CUBDL test set used in Chapter 3, we demonstrated that our improved method yields results comparable to networks specifically trained for speckle reduction tasks. This indicates the effectiveness and reliability of our approach in real-world scenarios. Furthermore, we conducted theoretical analyses to quantify the parameter overhead required for our proposed method. The results showed that the overhead is minimal, accounting for less than 0.07% in out implementation. This underscores the robustness and efficiency of our method, making it well-suited for practical applications.

5.2 Future Research

In this thesis, we have introduced two solutions for multitask learning applied to ultrasound beamforming. While deep learning-based beamformers have demonstrated superior results compared to classical algorithms, and specifically, multitask learning methods for lightweight multitask beamformers, several open questions for future re-

search emerge:

1. Scalability of Multitask Learning: In Chapter 3, we initially tested our model with two tasks. However, commercial ultrasound devices often incorporate more than two tasks, allowing for the combination of multiple post-processing filters. Therefore, it is imperative to investigate the scalability of our method for a larger number of tasks to accommodate the diverse functionalities of commercial ultrasound devices.

2. Generalization Across Ultrasound Probe Types: Our approach was tested on data obtained specifically from plane wave ultrasound. However, commercial ultrasound devices offer various types of probes, such as focused transmit or synthetic aperture, each employing different image formation pipelines. Consequently, it is essential to evaluate the proposed method's performance across different types of probes to ensure its versatility and applicability across diverse ultrasound imaging scenarios.

3. Evaluation of Real-Time Performance: Commercial ultrasound devices typically operate under stringent requirements for high frame rates, often aiming for around 30 frames per second. Although our proposed method utilizes a lightweight neural network, its performance on hardware similar to or identical to commercial ultrasound systems needs to be rigorously evaluated to determine its ability to meet the required frame rate demands. This assessment is crucial for assessing the practical viability of our proposed method in real-world clinical settings.

Bibliography

[BHH+20a] Muyinatu A Lediju Bell, Jiaqi Huang, Dongwoon Hyun, Yonina C Eldar, Ruud Van Sloun, and Massimo Mischi. Challenge on ultrasound beamforming with deep learning (cubdl). In *Proc. 2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–5. IEEE, 2020.

[BHH+20b] Muyinatu A. Lediju Bell, Jiaqi Huang, Dongwoon Hyun, Yonina C. Eldar, Ruud van Sloun, and Massimo Mischi. Challenge on ultrasound beamforming with deep learning (cubdl). In *Proc. 2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–5, 2020.

[BKH16] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[BNKL20] Manish Bhatt, Arun Asokan Nair, Kelley M. Kempski, and Muyinatu A. Lediju Bell. Multi-task learning for ultrasound image formation and segmentation directly from raw in vivo data. In *Proc. 2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2020.

[CFT12] Olivier Couture, Mathias Fink, and Mickael Tanter. Ultrasound contrast plane wave imaging. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 59(12):2676–2683, 2012.

[CHKB09] Pierrick Coupe, Pierre Hellier, Charles Kervrann, and Christian Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing*, 18(10):2221–2229, 2009.

[CRG+20] Aaron Carass, Snehashis Roy, Adrian Gherman, Jacob C Reinhold, Andrew Jesson, Tal Arbel, Oskar Maier, Heinz Handels, Mohsen Ghafoorian, Bram Platel, Ariel Birenbaum, Hayit Greenspan, Dzung L Pham, Ciprian M Crainiceanu, Peter A Calabresi, Jerry L Prince, William R Gray Roncal, Russell T Shinohara, and Ipek Oguz. Evaluating white matter lesion segmentations with refined Sørensen-Dice analysis. en. *Sci. Rep.*, 10(1):8242, May 2020.

[DSOL18] Fabian Dietrichson, Erik Smistad, Andreas Ostvik, and Lasse Lovstakken. Ultrasound speckle reduction using generative adversial networks. In

Proc. 2018 IEEE International Ultrasonics Symposium (IUS), pages 1–4, 2018.

[GAR20] Sobhan Goudarzi, Amir Asif, and Hassan Rivaz. Ultrasound beamforming using mobilenetv2. In *Proc. 2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2020.

[HB17] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. ICCV*, 2017.

[HG16] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[HG20] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2020. arXiv: `1606.08415 [cs.LG]`.

[HWG+21a] Dongwoon Hyun, Alycen Wiacek, Sobhan Goudarzi, Sven Rothlübbers, Amir Asif, Klaus Eickel, Yonina C. Eldar, Jiaqi Huang, Massimo Mischi, Hassan Rivaz, David Sinden, Ruud J. G. van Sloun, Hannah Strohm, and Muyinatu A. Lediju Bell. Deep learning for ultrasound image formation: cubdl evaluation framework and open datasets. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(12):3466–3483, 2021.

[HWG+21b] Dongwoon Hyun, Alycen Wiacek, Sobhan Goudarzi, Sven Rothlübbers, Amir Asif, Klaus Eickel, Yonina C Eldar, Jiaqi Huang, Massimo Mischi, Hassan Rivaz, et al. Deep learning for ultrasound image formation: cubdl evaluation framework and open datasets. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 68(12):3466–3483, 2021.

[IS15] Sergey Ioffe and Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proc. International conference on machine learning*, pages 448–456. pmlr, 2015.

[KHY21a] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye. Switchable and tunable deep beamformer using adaptive instance normalization for medical ultrasound. *IEEE Transactions on Medical Imaging*, 41(2):266–278, 2021.

[KHY21b] Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye. Switchable deep beamformer for ultrasound imaging using adain. In *Proc. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 677–680, 2021.

[KLA19] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[KNZL22] Hee Guan Khor, Guochen Ning, Xinran Zhang, and Hongen Liao. Ultrasound speckle reduction using wavelet-based generative adversarial network. *IEEE Journal of Biomedical and Health Informatics*, 26(7):3080–3091, 2022.

[LCdB+20] Ben Luijten, Regev Cohen, Frederik J. de Bruijn, Harold A. W. Schmeitz, Massimo Mischi, Yonina C. Eldar, and Ruud J. G. van Sloun. Adaptive ultrasound beamforming using deep learning. *IEEE Transactions on Medical Imaging*, 39(12):3967–3978, 2020.

[LH17] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. *CoRR*, abs/1711.05101, 2017. arXiv: `1711.05101`.

[LLY+21] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 2021.

[LRC+16] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J.A. Jensen, and O. Bernard. Plane-wave imaging challenge in medical ultrasound. In *Proc. 2016 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2016.

[LZ20] Yancheng Lan and Xuming Zhang. Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet. *IEEE Access*, 8:195327–195340, 2020.

[MF09] Juan L Mateo and Antonio Fernández-Caballero. Finding out general tendencies in speckle noise reduction in ultrasound images. *Expert systems with applications*, 36(4):7786–7797, 2009.

[MMP+19] Moein Mozaffarzadeh, Ali Mahloojifar, Vijitha Periyasamy, Manojit Pramanik, and Mahdi Orooji. Eigenspace-based minimum variance combined with delay multiply and sum beamformer: application to linear-array photoacoustic imaging. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–8, 2019.

[PGC+17] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch, 2017.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing, 2015.

[RSE+20] Sven Rothlübbers, Hannah Strohm, Klaus Eickel, Jürgen Jenne, Vincent Kuhlen, David Sinden, and Matthias Günther. Improving image quality of single plane wave ultrasound via deep learning based channel compounding. In *Proc. 2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2020.

[SG17] Sameera V Mohd Sagheer and Sudhish N George. Denoising of medical ultrasound images based on non-local similarity: a low-rank approach. In *Proc. TENCON 2017 - 2017 IEEE Region 10 Conference*, pages 176–181, 2017.

[SHZ+18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: inverted residuals and linear bottlenecks. In *Proc. of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[SK16] Tim Salimans and Durk P Kingma. Weight normalization: a simple reparameterization to accelerate training of deep neural networks. *Advances in neural information processing systems*, 29, 2016.

[WSB03] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *Proc. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, 1398–1402 Vol.2, 2003.

[XCZ+20] Ruoxiu Xiao, Cheng Chen, Hanying Zou, Ying Luo, Jiayu Wang, Muxi Zha, and Ming-An Yu. Segmentation of cerebrovascular anatomy from TOF-MRA using length-strained enhancement and random walker. en. *Biomed Res. Int.*, 2020:9347215, September 2020.

שנלמדו על ידי רשת נוירונים בשכבות השונות. על ידי שינוי הייצוגים ניתן להסיט את הידע של הרשת בין משימות שונות. גם במסגרת חלק זה ערכנו את הבדיקות שהוצעו במסגרת האתר ליצירת תמונת אולטראסאונד מבוסס למידה עמוקה. בחלק זה השווינו את הרשת שלנו לגישות קלאסיות (ללא למידה עמוקה) וגם לרשת שאומנה ספציפית עבור משימה של ניכוי רעשי מריכה.

בפרק זה השיטה שלנו הציגה תוצאות טובות באופן משמעותי מן הגישות הקלאסיות, בנוסף, הראנו כי התוצאות קרובות מאוד (עד כדי הבדל זניח) לרשת שאומנה עבור משימת היעד באופן ספציפי. עובדה זו מוכיחה את אפקטיביות הגישה המוצעת ללמידה מרובת משימות. לבסוף, הצגנו ניתוח תאורטי של כמות הפרמטרים הנלמדים הדרושים עבור כל משימה עבור הגישה שלנו. הראנו כי במקרים פרקטיים סך הפרמטרים הנדרשים עבור למידה של משימה נוספת הוא פחות מאחוז מכמות הפרמטרים המקורית של הרשת – עובדה שמצביע על כך שהשיטה המוצעת גם יעילה מבחינת זמן חישוב ומקום. לאור כל הבדיקות שביצענו אנו מסיקים כי הגישות שהוצגו ניתנות לשימוש במערכת אולטרסאונד שכן הן מאפשרות לימוד של מספר רב של משימות, עם תקורת פרמטרים מינימלית ותוצאות ויזואליות שמנצחות את הגישות הקלאסיות ועמדות בשורה אחת עם רשתות שאומנו למשימות באופן ייעודי.

תקציר

בחיבור זה, אנו חוקרים את אתגרי עיצוב עלומות למערכות סריקת אולטראסאונד רב-משימות. עיצוב אלומות הוא טכניקה קריטית לשיפור איכות האות מכיוונים ספציפיים תוך הפחתת הפרעות ממקורות אחרים. נעשה שימוש בטכניקה זו במגוון שיטות דימות רפואי ובמיוחד בסריקות אולטרסוניות, שבהן היא משפיעה משמעותית על רזולוציית התמונה והבהירות.

התיזה עוסקת בשני בעיות עיקריות. הבעיה הראשונה מתמקדת בפיתוח של שיטת למידת רב משימתית מבוססת על נורמלזיציה של משקולות גרעין קונבולציה עבור יצירת תמונת אולטראסאונד. המטרה היא להציע מודל אחד שיכול לשחזר תמונה מאות תת דגום ולבצע הפחתת רעשים באופן שנשלט על ידי המשתמש. על ידי התאמת פרמטרים על ידי שיטת הנומליזציה המוצעת, למשימה של שיחזור מאות תת דגום, ואופטימיזיה שלהם להפחתת רעשי מריכה, רשת הלמידה העמוקה השיגה ביצועים טובים יותר משיטות היצירת תמונת אולטראסאונד המסורתיות במדדי השגיאה המקובלים בתחום. הבעיה השנייה חוקרת גישת למידה עמוקה קלת משקל המבוססת על ריבוי משימות ליצירת תמונת אולטרסאונד. גישה זו שואפת לשפר את הרזולוציה והניגודיות של התמונה, הנמוכים בדרך כלל באולטרסאונד בהשוואה לשיטות הדמיה אחרות. אנו מציגים שיטה חדשה שלומדת טרנספורמציה של ייצוגים ממשימת מקור של יצירת תמונת אולטראסאונד למשימת יעד של הפחתת כתמים בצורה יעילה בפרמטרים, מה שמבטיח עלויות חישוב נמוכות. הערכות איכותיות מוכיחות ששיטה זו משיגה ביצועים הדומים לאימון רשת ייעודית תוך ביצועים גבוהים יותר משמעותית משל טכניקות השהייה וסכום בהפחתת רעש כתמים.

ראשית, אנו מציגים את השיטה הראשונה שלנו ללמידה מרובת משימות. שיטה זו מבוססת על גישה ייחודית של התאמת משקולות ממשימה מסוימת למשימות אחרות. ההתאמה נעשית על ידי למידת פרמטרי נרמול עבור כל שכבה ברשת הקונבלוציה שאומנה למשימת בסיס. גישה זו מאפשר התאמה של ידע של רשת ממשימת בסיס, שבמקרה שלנו היא יצירת תמונת אולטרסאונד, למשימות יעד שונות. בפרק זה בחרנו לבדוק את הגישה שלנו על שתי משימות נוספות: שיחזור מאות תת דגום וניקוי כתמי מריחה. בפרק זה ערכנו שני שלבי בדיקה. בשלב הראשון בדקנו את הרשת שאומנה עבור משימת הבסיס השווינו אותה למשתתפים אחרים באתגר יצירת תמונה מאולטרסאונד על ידי למידה עמוקה. המודל שלנו השיג את התוצאות הגבוהות ביותר בכל המדדים למעט מדד אחד. לאחר מכן בדקנו את אפקטיביות הגישה שלנו בלמידה מרובת משימות. גם בחלק זה השתמשנו במדדים שהוצגו במסגרת האתגר והשווינו את המודל שלנו לשיטות מקבילות. גם בחלק זה המודל שלנו הציג את התוצאות הגבוהות ביותר. באופן ספציפי, הניגודיות ומדד הרעש של התמונות ששוחזרו בעזרת השיטה שלנו היו טובים באופן ניכר מן המתחרים. לאחר מכן, הצגנו שיטה נוספת ללמידה עמוקה מרובת משימות. במסגרת חלק זה פיתחנו שיטה חדשה ללמידה יעילה ודלת משאבים של שינוי ייצוגים

המחקר בוצע בהנחייתו של פרופסור ישראל כהן בפקולטה להנדסת חשמל ומחשבים.

התוצאות של חיבור זה פורסמו כמאמר מאת המחבר ושותפיו למחקר בכתב-עת במהלך תקופת מחקר המגיסטר של המחבר.

מחבר חיבור זה מצהיר כי המחקר, כולל איסוף הנתונים, עיבודם והצגתם, התייחסות והשוואה למחקרים קודמים וכו', נעשה כולו בצורה ישרה, כמצופה ממחקר מדעי המבוצע לפי אמות המידה האתיות של העולם האקדמי. כמו כן, הדיווח על המחקר ותוצאותיו בחיבור זה נעשה בצורה ישרה ומלאה, לפי אותן אמות מידה.

תודות

ברצוני להודות למנחה המחקר שלי, פרופ' ישראל כהן, שלולא תמיכתו ותרומתו מחקר זה לא היה מתאפשר. לאורך הפרויקט הזה הוא הדריך אותי כיצד לבצע מחקר ובעזרתו רכשתי מיומנויות יקרות ערך שעזרו לי להתגבר על קשיים ופערי ידע בתחומים מסוימים. ההדרכה שלו הייתה מכרעת בהתפתחות שלי במהלך המאמץ המחקרי הזה. בנוסף, אני רוצה להודות לחבריי לעבודה על ההתייעצויות והתרומה לעבודה זו.

אני רוצה להודות גם להורי חיים וסיגל ולבת זוגתי נועה. ללא תמיכתם לאורך המסע הזה, המחקר הזה לא היה מתאפשר.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

למידה עמוקה מרובת משימות לעיצוב אלומת אולטרסאונד

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר
מגיסטר למדעים בהנדסת חשמל

עיליי דהן

הוגש לסנט הטכניון — מכון טכנולוגי לישראל
שבט התשפ"ה חיפה פברואר 2025

למידה עמוקה מרובת משימות לעיצוב אלומת אולטרסאונד

עיליי דהן